Análise visual do Twitter

Visual analysis of Twitter

Elias Estevão Goulart

Doutor pela Universidade de São Paulo e Pós-Doutor pela University of British Columbia, Canadá. Professor e pesquisador do Programa de Pós-Graduação em Comunicação da Universidade Municipal de São Caetano do Sul. Coordenador Adjunto do Laboratório de Hipermídias. elias.goulart@uscs.edu.br.

Sidney Fels

Doutor pela University of Toronto. Professor do Departamento de Engenharia Elétrica e Computação da University of British Columbia. Líder do Human Communication Technologies Laboratory. ssfels@ece.ubc.ca.

Recebido em 10 de junho de 2014. Aprovado em 22 de setembro de 2014

Resumo

A mídia social tornou-se um canal muito importante de comunicação e interação para pessoas de todo o mundo, e uma grande quantidade de conteúdo está sendo criado. Como resultado, o processo de análise de tal enorme quantidade de dados requer o suporte de ferramentas e técnicas de visualização. Este estudo está centrado nas relações entre as palavras postadas no Twitter, usando o princípio da Folksonomia para categorizar as palavras mais recorrentes como etiquetas (ou tags). Além disso, ele propõe um modelo visual baseado no princípio de atração física, que tem como objetivo mostrar a maneira como as principais etiquetas estão correlacionadas. Os resultados indicam o potencial do Modelo Orbital, porque pode ser utilizado para representar a dinâmica das relações ao longo do tempo.

Palavras-chave: Comunicação, Inovação, Tecnologias Digitais, Mídias Sociais, Twitter, Folksonomia, Visualização de Dados.

Abstract

Social media has become a very important channel of communication and interaction for people all over the world and a large amount of content is being created. As a result, the process of analyzing such huge amount of data requires the support of visualization tools and techniques. This study focuses on the relationships between the words posted on Twitter, using the Folksonomia principle to categorize the most recurrent words as tags. Additionally, it proposes a visual model based on the physical attraction principle that aims to show the way that the main tags are correlated. The results indicate the potential of the orbital model since it can be used to represent the dynamics of relationships over time.

Keywords: Communicataion, Inovation, Digital Technologies, Social Media, Twitter, Folksonomy, Data Visualization.

Introdução

As redes sociais virtuais (RSV) são fenômenos contemporâneos de grande importância para a compreensão das relações entre as tecnologias digitais e as mudanças que causam nos hábitos e comportamentos das pessoas. A grande população de usuários, que cresce de forma dramática, e a enorme quantidade de conteúdo gerado por eles criam um banco de dados planetário on-line que pode ajudar a compreender as relações interpessoais em todas as dimensões humanas, tais como educação, saúde, lazer, segurança etc. (Bogutá, 2009).

Este grande volume de informações requer cuidados especiais, coleta e agregação, análise e inferência e formas de apresentação e visualização, para que se possa chegar a uma interpretação e à compreensão das possíveis relações e significados embutidos nelas. Vários estudos em RSVs indicaram este desafio como um verdadeiro obstáculo para o desenvolvimento de soluções em sistemas de informação, ou seja, não apenas as exigências de projeto e construção desses novos sistemas relacionados com a organização de dados e modos de operação nas redes sociais virtuais, mas como empregá-los utilmente e o que se pode fazer por meio deles. Tem sido dito que uma grande quantidade de informação fala por si mesma (Boyd & Crawford, 2012).

A representação gráfica de dados abstratos não é nova, uma vez que tem sido usada na computação desde a criação de interfaces gráficas com base no mecanismo de janelas. Portanto, a visualização de informações é um aspecto importante da computação gráfica e tem subsidiado a construção de sistemas para monitorar o conteúdo e representação de redes sociais virtuais.

O foco deste estudo é analisar o conteúdo do Twitter, tentando fornecer uma apresentação gráfica de seus *posts* e mostrar as possíveis relações entre seu conteúdo, de preferência através de visualizações dinâmicas.

Estudos sobre o conteúdo do Twitter seguem, principalmente, três linhas diferentes:

- A análise dos nós da rede e ligações entre usuários, suas topologias e distribuição, com base principalmente na teoria dos grafos (Chen & Yang, 2010; Skold, 2008);
- Estudos com foco no engajamento e influência dos utilizadores, tendo em conta a quantidade de seguidores, responde a mensagens (*replys*), a republicação de mensagens (*retweets*), citações específicas sobre questões e/ou utilizadores (*hashtagging*), bem como abordagens para perfis de usuários, sua geolocalização etc. (himelboin, Hansen & Bowser, de 2012; Becker, Naaman & Gravano, 2011; Romero, Meeder & Kleinberg, 2011);

– Avaliação do conteúdo, com foco em linguística, ou em estudos de Processamento de Linguagem Natural, para identificar questões discutidas e/ou aspectos de usuários, tais como sentimentos, atitudes e percepções, com base em listas de palavras ou frases-chave (Pang & Lee, 2008; Liu, 2010; Tsytsauru & Palpanas, 2011; Go et al., 2009; Zhang et al., 2011).

Assim, o objetivo deste trabalho é apresentar um modelo de representação visual que pode revelar relações entre as palavras usadas por usuários do Twitter quando falam sobre um determinado assunto. Neste caso, os comentários expressam suas opiniões, ideias, críticas etc., e as palavras usadas com mais frequência são as que melhor expressam as mesmas ideias desses usuários. Oportunamente, outro estudo mais específico poderá discutir as possíveis aplicações deste modelo e suas potencialidades e limitações.

Em geral, uma ferramenta visual que mostra a relação entre um determinado assunto e as principais palavras-chave utilizadas pode indicar tendências de opiniões e expressões dos usuários em um determinado momento. Além disso, se o modelo puder mostrar a dinâmica entre o sujeito e as palavras usadas ao longo do tempo, ou seja, se houver uma indicação de fortalecimento ou enfraquecimento da relação, isso permitirá uma melhor compreensão das relações e seus significados.

Este estudo traz uma contribuição sobre o uso da Folksonomia para priorizar a seleção de conteúdos do Twitter, bem como sobre o uso de uma analogia com um princípio da física para estudar e representar visualmente a relação entre um determinado assunto e os termos usados em mensagens na maioria das vezes, resultando em uma exposição gráfica das tendências dinâmicas para a sua expressão.

Por fim, continua com uma discussão sobre as formas de representação visual de informações do Twitter, a apresentação do conceito da Folksonomia e seu uso associado à análise de conteúdo, e com o modelo de representação proposto para o acompanhamento visual de informações a partir de redes sociais virtuais.

Twitter

O Twitter é uma plataforma de computação baseada na Web que foi construída para suportar comunicações curtas e rápidas. Os usuários a empregam para as mais diversas finalidades, tais como relatar eventos, acontecimentos, opiniões, ideias, notícias e informações multimídia a serem compartilhadas com a família, amigos, alunos etc., (Cheong & Lee, 2010). Também conhecido como um *microblog*, a plataforma se

diferencia dos blogs convencionais pelo pequeno tamanho de suas mensagens (frases publicadas por usuários com o comprimento máximo de 140 caracteres), o que facilita a produção de textos e não requer muito tempo para a criação intelectual em profundidade, tornando a publicação mais ágil e com maior produção de mensagens (geralmente várias vezes por dia) para um usuário comum (Java et al., 2009).

Sua popularidade mundial trouxe uma nova demanda: sua utilização como um canal de comunicação com abrangência e cobertura de questões de relevância social, como campanhas eleitorais (EUA, Egito etc.), ações de promoção da saúde, monitoramento de tragédias, acompanhamento de prisioneiros políticos (por exemplo, jornalistas), ataques terroristas, entre outros (Fleishman, 2009; Jungherr, 2009; Goolsby, 2009). Assim, o Twitter é mais do que uma RSV que simplesmente permite o estabelecimento de relações simples entre pessoas seguidoras e seguidas, mas também se constitui em um universo de conexões para interações quase instantâneas, inteiramente adequadas para a mineração e visualização de dados.

Neste sentido, a rede social virtual tem uma aplicação específica como um canal de comunicação em que as pessoas podem expressar livremente suas opiniões, ideias, reclamações etc., principalmente para seus seguidores, mas também para o público em geral. Alguns estudos recentes analisaram seu uso como uma forma de acompanhar as postagens dos usuários sobre marcas, como uma pesquisa de mercado em tempo real (Merles World, 2012;. Jansen et al, 2009; Nicholls, 2012; Pang & Lee, 2008; Thelmall, Buckley & Paltoglou, 2011).

A questão que se coloca é como "ver" ou representar visualmente a presença ou a importância de termos ou palavras-chave que são incorporados a um grande número de postagens, de forma a trazer significados específicos para o analista – um pesquisador, o proprietário de uma marca comercial ou um simples usuário.

Parte dos estudos sobre o Twitter enfocam a visualização de dados das conexões entre usuários, sua densidade e morfologia. Esta área tem sido chamada de Análise de Redes Sociais, a qual inclui a verificação de *clusters*, nós centrais, dentre outras estruturas, sua centralidade, capilaridade e vários outros recursos usados para comparar e classificá-los, como fizeram Shold (2008), Lin e Dyer (2010) e Highfield, Kirchhoff e Nicolai (2011).

Outros estudos estão focados no tratamento de variáveis quantitativas, como o número de seguidores, número de postos de trabalho, a ocorrência de *retweets* e o conteúdo específico como etiquetas (tags) especiais, as *hashtags*, ou *links* nas mensagens. Estes dados permitem análise estatística, correlações e vários gráficos para apresentá-los, como nos trabalhos de Gilbert, Karahalios e Sandvig (2010), Becker, Naaman e Gravano (2011) e Thelmall, Paltoglou e Buckley (2011).

As representações também podem ser construídas para comparar as variáveis qualitativas ou categóricas com base nos termos (palavras-chave) utilizados por tais representações como gráficos de barras e gráficos de pizza, entre outros. Nestes casos, a ideia consiste em comparar a forma como estes elementos são apresentados na associação das informações recolhidas. Na verdade, o tipo de apresentação depende das categorias de informações, tais como o impacto das mensagens de usuários, o seu grau de influência, de engajamento etc., como nos estudos de Cheong e Lee (2010) e Kawano, Kishimoto e Yonekura (2011). Alguns sites oferecem representações de informações do Twitter: o Twitalyzer (www.twitalyzer.com), o Twitter Contador (www.twittercounter.com) Twitter Stats (www.twitterstats.com), o We Feel Fine Projeto (wefeelfine.org), dentre outros.

O monitoramento da ocorrência de termos em redes sociais virtuais é uma ferramenta de avaliação importante para o marketing, a política, a educação e outros campos, revelando os desejos, as necessidades, as preocupações e as posições das pessoas ligadas a eles. O método que subsidia esses estudos é a Análise de Conteúdo, que permite o estabelecimento de categorias, resultando em palavras-chave de interesse para pesquisar as mensagens (Wilson, 2011). A definição de categorias não é uma tarefa trivial, especialmente quando não se tem, *a priori*, uma base teórica para a terminologia. Portanto, a forma mais adequada de fazer isso é por meio da observação da ocorrência de palavras, a fim de destacar as questões e as palavras-chave associadas a eles. Esta mineração de informações pode ser complexa, dependendo da coleção disponível e, geralmente, requer algum auxílio computacional.

Neste estudo, em particular, a observação é mais desafiadora pelo dinamismo das mensagens, bem como pela alternância de temas e das pessoas que falam sobre eles, já que não há controle sobre esse enorme ambiente.

Folksonomia

O nome Folksonomia vem das palavras "taxonomia" e "folclore" (popular). "Taxonomia" representa uma classificação conceitual criada por especialistas e "popular" refere-se a um grupo de pessoas que produz algum tipo de ação ou de cultura. Portanto, a Folksonomia indica uma estrutura de classificação criada por um grupo de pessoas que não são especialistas (Schmitz, 2006).

Thomas Vander Wal, em 2004, sugeriu esse termo para questionar se o possível desenvolvimento de uma estrutura de categorização criada por usuários, em um modo informal e através de uma abordagem "*bottom-up*", poderia levar a um *thesaurus* emergente. Para formular essa questão, ele usou a palavra "Folksonomia" (López-Juárez

& Olivas, 2011). Vander Wal comentou que essa forma de classificação deve incluir o recurso a ser classificado, o código de classificação (tag) e a identidade do classificador. Esta tríade é a base para a construção do sistema de classificação, atualmente disponível em muitos sites que permitem o registo e armazenamento on-line de "recursos", como fotos (Flicker), vídeos (YouTube), texto (Blogger), *links* para páginas da Web (Delicious) etc.

Como exemplo, um usuário pode armazenar uma foto em um site e adicionar a etiqueta que, em sua opinião, melhor descreve a foto. Outro usuário, que encontra a mesma imagem e está interessado nela, poderá usar a mesma etiqueta, que já tinha sido associada a ela ou pode escolher outra etiqueta mais adequada para o seu próprio uso, a fim de facilitar sua posterior recuperação. Assim, com muitos usuários acessando essa imagem ao longo do tempo, havia uma ou mais etiquetas que foram mais utilizadas para designar ou descrever aquela foto em especial e, portanto, essas são as etiquetas que melhor a representam, de acordo com o princípio da Folksonomia.

A Folksonomia tem sido utilizada para diversos fins em sistemas que fornecem o mecanismo para a rotulagem (etiquetagem), a fim de criar recursos de informação comumente usados (*ranking*), como a determinação de pontuação de *links* (Jin et al. 2011), a organização de informações em bibliotecas (Ran & Erpeng, 2011), os sistemas de recomendação para a apresentação dos resultados de pesquisas (Luo, Ouyang & Xiong, 2012; Kim et al., 2012), a criação de ontologias para auxiliar os processos de ensino e aprendizagem (Petrucco, 2011), entre muitos outros.

Além disso, incentiva o engajamento social em uma comunidade de pessoas com interesses semelhantes por meio do processo criativo de elaboração das etiquetas com base em um vocabulário comum, muitas vezes revelando um entendimento comum de um modelo conceitual, ao contrário de uma taxonomia concebida como um modelo hierárquico de categorização do conhecimento por especialistas (Kakali & Papatheodorou, 2010).

Neste estudo, o mecanismo de seleção de etiquetas mais frequentes será aplicado para criar uma lista das etiquetas mais usadas no Twitter, o que significa que cada termo ou palavra usada nas mensagens será considerada como etiqueta escolhida pelos usuários para descrever melhor ou para indicar o sentido do que querem expressar. Em outras palavras, a fim de analisar o conteúdo das mensagens e estabelecer possíveis relações entre as categorias ou palavras-chave, o conceito de etiquetagem será empregado para todas as palavras usadas nas mensagens. O objetivo é verificar se há uma escolha mais adequada ou mais um indicativo de palavras por parte dos usuários para se referir a um determinado assunto ao usar o Twitter.

Modelo Orbital aplicado ao Twitter

O modelo considera um grupo de mensagens que fala sobre um assunto específico (expresso por uma palavra-chave) e cada mensagem é composta por palavras, com um máximo de 140 caracteres, que é a estrutura do Twitter, de onde foram coletadas as mensagens.

As palavras mais utilizadas seriam consideradas as mais adequadas, de acordo com os usuários que optaram por elas. Aprender sobre como essas palavras estão relacionadas com o tema da discussão é algo que pode ser representado graficamente por meio do princípio da atração física. Dessa forma, os termos (ou palavras) teriam uma "atração" hipotética para com a palavra-chave (que representa o assunto) expressa pelo número de vezes que aparecem nas postagens coletadas. Além disso, o seu significado é determinado pelo contexto em que foram publicadas as mensagens, por exemplo, os comentários dos usuários sobre um político são considerados no contexto de uma determinada campanha eleitoral, ou uma recordação de um determinado veículo, no contexto de um determinado anúncio fabricante, bem como a língua, a cultura ou outros aspectos relacionados com os usuários. Por estes exemplos, as palavras mais usadas nas postagens sobre estes assuntos são as que melhor expressam as opiniões dos usuários.

A identificação de palavras, ou etiquetas, como o conceito de Folksonomia estabelece, vem da análise de conteúdo, que separa as palavras em uma lista e indica a contagem das ocorrências de cada palavra em cada postagem. O assunto em si, o foco da análise, será representado por uma palavra-chave, o que em teoria deveria ser a etiqueta mais recorrente.

O princípio da atração física (a força gravitacional, ou a força elétrica) postula que a atratividade é diretamente proporcional à magnitude característica dos objetos (quantidade de massa no caso gravitacional ou a quantidade de carga, se elétrico), inversamente proporcional ao quadrado da distância entre eles e, de forma proporcional, dependente do ambiente, representada por uma constante, ou seja:

onde:
F = força de atração
K = constante associada ao ambiente
 = magnitude dos objetos
D = distância entre os objetos

A lista produzida pela análise folksonômica das postagens do Twitter apresenta, em ordem decrescente de número de repetições, as palavras que foram mais utilizadas

pelos usuários, dado um número N de palavras coletadas em um determinado período de tempo. A palavra-chave que representa o assunto em questão aparecerá em primeiro lugar na lista com um número de ocorrências P (menor ou igual a N). A segunda palavra na lista, que será a etiqueta mais usada, terá uma quantidade S e estará à distância de uma unidade (valor 1); a terceira palavra na lista, o que significa a segunda palavra mais amplamente utilizada, terá uma quantidade T e estará na distância de duas unidades, ou segunda da lista, (valor de 2), e assim por diante. Note-se que os valores de P, S e T podem ser iguais a N se as palavras aparecem em todas as postagens. Ainda, o valor de P é considerado como referência porque é o centro das conversas sobre as mensagens.

A constante K depende da natureza do ambiente do fenômeno observado. Neste estudo, estamos operando com palavras ou, mais genericamente, com os códigos em uma linguagem particular. Portanto, é necessário levar em consideração a possibilidade de ocorrência dos códigos e o número de códigos disponíveis para representar certo "discurso" ou expressão. Além disso, é necessário normalizar os resultados, a fim de ter um modelo capaz de ser usado em "qualquer língua", especialmente quando se quer comparar os resultados obtidos em um determinado assunto em diferentes idiomas.

Assim, este estudo sugere que a constante K deve ser a relação entre o número de palavras originais usados pelos usuários, dividido pelo total de ocorrências em que todas as palavras das mensagens, reduzindo o efeito das diferenças entre os conjuntos de códigos em diferentes idiomas. Isso acontece porque o resultado está vinculado aos dados que estão sendo analisados em um determinado momento, o que limita a um conjunto de palavras empregadas pelos usuários sobre o assunto em foco.

Como um resultado desta modelagem, pode-se criar uma representação visual, tal como a ilustrada na Figura 1, em que um conjunto particular de etiquetas relacionadas com a palavra-chave (ou assunto) pode ser visualizado, bem como as suas dimensões em relação à palavra-chave em um determinado tempo. Além disso, por meio da recolha de mais postagens ao longo do tempo será possível verificar a dinâmica da relação entre as palavras, ou seja, se mais usuários estão "dizendo" as mesmas etiquetas sobre o assunto, aumentando seu tamanho e importância (forma). Além disso, a aproximação ou o distanciamento das etiquetas em relação à palavra-chave, ou seja, a variação da posição da etiqueta na lista original e no gráfico visual pode indicar a sua prioridade sobre todas as outras.

Ao determinar a força de atração entre o assunto e as etiquetas relacionadas, podemos comparar, por exemplo, a relação entre os comentários dos usuários sobre o assunto

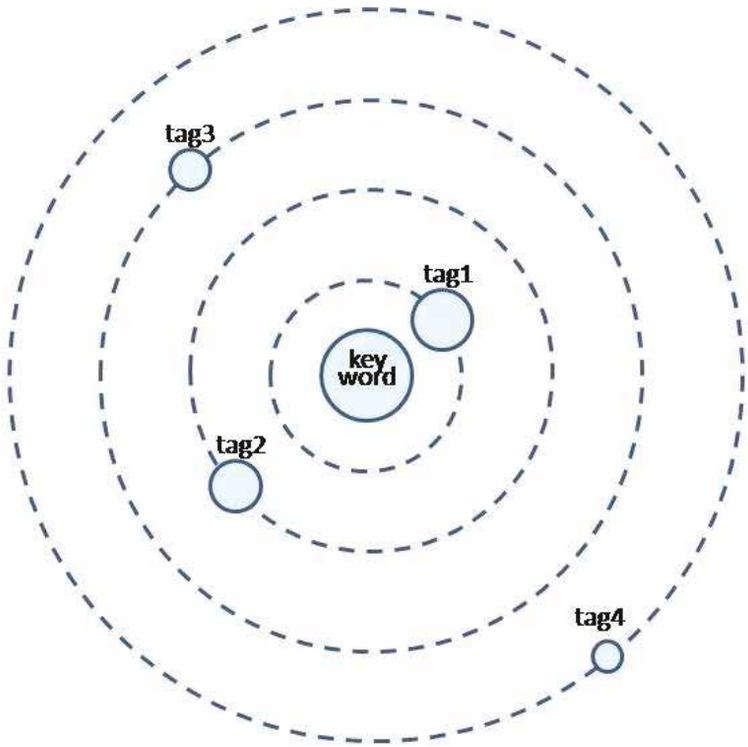

Figura 1: Representação do Assunto e as Etiquetas relacionadas.

em diferentes localidades, pois o cálculo aqui proposto leva em conta o contexto linguístico das postagens.

O gráfico é construído de forma relativa: o círculo central que representa a palavra-chave terá o valor "100" (número de referência) em relação ao número de postagens obtidos com a palavra-chave. O tamanho dos círculos, que representam as outras etiquetas, será determinado proporcionalmente à palavra-chave (número de referência) com base no número de suas ocorrências. Supondo que todas as etiquetas sejam mencionadas dentro de cada postagem, assim como a palavra-chave, todos os "planetas" no gráfico terão o mesmo tamanho, como mostrado na figura 2.

Desta forma, é possível observar dois aspectos importantes:

a) A mudança de objetos entre órbitas indica uma aproximação (ou distanciamento) de uma etiqueta em relação à palavra-chave escolhida. A mudança entre órbitas indica um aumento (ou diminuição) da importância da relação dos termos (palavra-chave e etiqueta) ou sua atratividade. Isto é conseguido pela alteração da posição relativa de uma etiqueta na lista ordenada de palavras descarregadas do Twitter;

b) Em segundo lugar, o raio do círculo que representa uma etiqueta está associado a sua contagem de ocorrências. A etiqueta pode ser visualmente perceptível pela dimensão

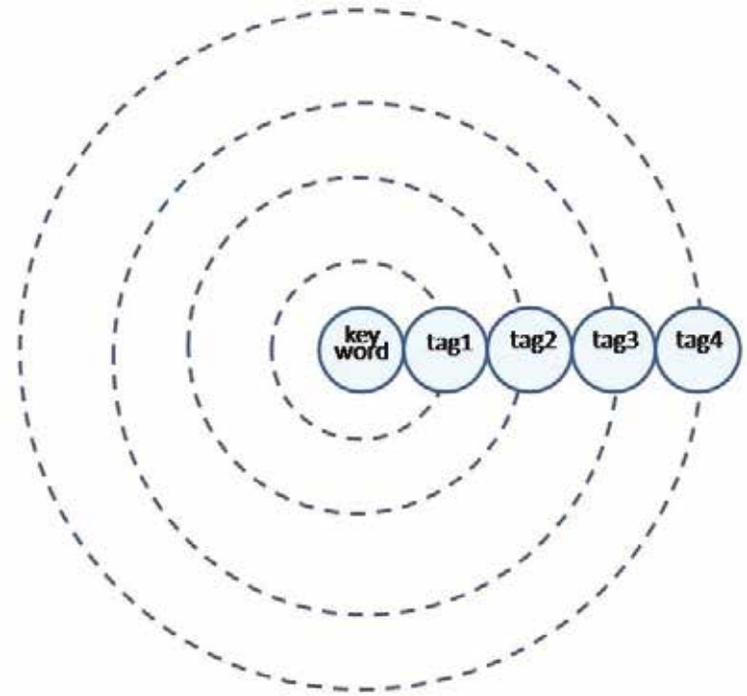

Figura 2: Representação de etiquetas com mesma ocorrência (tamanhos iguais).

do círculo. Assim, se a sua dimensão aumenta (em uma representação bidimensional) e manteve-se na mesma órbita, isso significa que há mais citações com esse termo e sua importância em relação à palavra-chave também aumenta (ou diminui).

A coleta e geração do banco de dados de mensagens foram realizados por meio de um aplicativo desenvolvido com a linguagem TCL (*Tool Command Language*), chamado "Trawler" (rastreador para Twitter), que utiliza as interfaces de programação SEARCH e REST (*Representational State Transfer*) fornecido pelo Twitter.

Exemplo de aplicação

A fim de ilustrar a aplicação do modelo, foi utilizado o nome da atual presidente do Brasil, Dilma, e uma busca foi feita para encontrar as mensagens que incluíam seu nome. Por padrão, o Twitter envia até 1.500 mensagens postadas recentemente (até 20 dias) quando se utiliza o comando SEARCH. Assim, como mostrado na Tabela 1, uma lista de palavras associadas foi construída, em que a palavra DILMA apareceu em quase todas as 100 postagens baixadas. Houve várias outras palavras retiradas da análise, pois eram apenas conectores (artigos, preposições) e não possuíam qualquer significado específico. Este procedimento mostra, mais uma vez, a necessidade de processar as

informações antes da análise, como indicado por vários outros estudos (Cheong & Lee, 2011; Cantadora, Konstas & Josec, 2011; Kakali & Papatheodorou, 2010), embora, este caso seja mais simples em termos da linguística, pois somente se deve compilar uma lista específica de conectores e a ferramenta irá excluí-los da lista original. Desta forma, a tabela 1 mostra as palavras selecionadas que foram ordenadas pelo número de sua ocorrência dentro de mensagens.

Tabela 1: Postagens relacionadas com a palavra-chave DILMA

	TERMO	Quantidade de ocorrências	
		02/05/12	04/05/12
0	DILMA	98	100
1	GREVE	9	13
2	FEDERAIS	7	9
3	HOMOFOBIA	7	3

Após o procedimento inicial, a lista com as 20 palavras mais frequentes resultou em apenas três etiquetas selecionadas com um significado específico para a análise, pois todos os outros eram conectores. A primeira a aparecer foi GREVE ("strike", em Inglês), pois no momento da pesquisa (2 de maio de 2012) a greve da Polícia Federal brasileira era um tema amplamente discutido no Brasil (veja que a palavra "Federal" foi a segunda da lista). Note-se que, dependendo da necessidade ou interesse da investigação, alguns conectores devem ser considerados, por exemplo, a palavra NÃO ("no" em Inglês) apareceu 78 vezes, mas é genérica e deve ser associada com outras palavras para ser considerada. A palavra "RT" também apareceu na lista original e significa "retuitar" no próprio vocabulário do Twitter, ou seja, uma postagem republicada de alguém que é seguido pelo editor; neste exemplo, elas foram descartadas. O texto de Manish Gupta et al. (2011) apresenta uma visão abrangente da sistematização do processo de etiquetagem.

Nesta coleção de postagens, 1.463 palavras apareceram no total, entre as quais 637 eram originais, o que torna o valor da constante K igual a 0,435, expressando a probabilidade de uma citação de cada palavra nessas postagens. A palavra-chave DILMA tinha apenas 98 ocorrências, assim o valor foi normalizado para 100, mudando a palavra GREVE para 9.184. Assim, a força de atração entre as palavras foi:

$$\rightarrow \quad F = 399{,}5$$

É importante notar que o valor máximo de F será 10.000 que, hipoteticamente, ocorre quando K = 1 e a primeira palavra na lista é apresentada em todas as mensagens

analisadas. O raio, ou a distância entre os termos, é 1, porque a palavra GREVE é a segunda na lista. A avaliação da força de atração entre "Dilma" e "Federais" seria igual a:

$$\rightarrow \quad F = 77{,}6$$

Outra coleta de dados foi feita dois dias após a primeira (4 de maio de 2012) e a palavra GREVE ainda estava no topo da lista, aparecendo com 13 ocorrências, enquanto a palavra-chave DILMA apareceu em todas as 100 postagens baixadas (Tabela 1). O novo cálculo mostra um aumento na força da relação, como pode ser visto por:

$$\rightarrow \quad F = 565{,}5$$

A Tabela 2 mostra os valores de atratividade entre palavra-chave e etiquetas. Aqui observa-se um aumento de 41,5%, o que pode ser interpretado como a importância do assunto.

Tabela 2: Variação da força de atração

TERMOS	02/05/12	04/05/12	Variação (%)
GREVE	399,5	565,5	+ 41,5
FEDERAIS	77,6	97,8	+ 26,0
HOMOFOBIA	34,5	14,5	- 57,9

O valor de 13 menções à palavra GREVE, em comparação com o dia anterior, será representado visualmente com um círculo de quase duas vezes o diâmetro do dia anterior, com um vetor colocado a seu lado, descrevendo o valor de variação e indicando o fortalecimento dessa questão entre os brasileiros usuários do Twitter nesses dias. É importante ressaltar que o conteúdo do Twitter é muito dinâmico e alguns casos têm alto impacto sobre os meios de comunicação, podendo fazer as postagens mudarem muito rapidamente, por isso é uma ferramenta com poder de mostrar visualmente esses fenômenos, contribuindo, assim, para seu entendimento. A Figura 3 mostra a dinâmica dessa relação.

Além disso, a determinação e monitoração do valor da força F, bem como a sua representação gráfica, tem o potencial de indicar a forma como está sendo discutido determinado assunto por usuários naquele momento e pode servir como uma comparação para verificar a sua "evolução" ao longo do tempo.

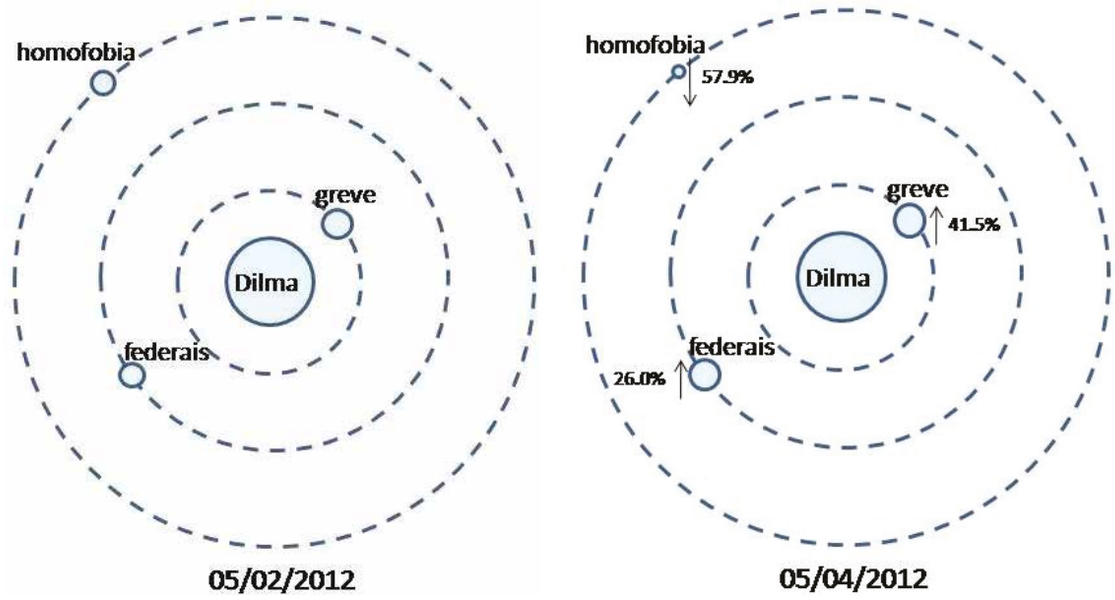

Figura 3: Representação visual da dinâmica da relação entre as palavras.

Este exemplo mostra o potencial do Modelo Orbital para representar visualmente as ocorrências de argumentos sobre um determinado assunto de interesse e permitindo o acompanhamento do progresso das discussões ao longo do tempo.

Considerações Finais

A representação visual de informações é uma característica importante e tem sido sempre um desafio para o design de interfaces de sistemas de computação. A exibição de dados gráficos e de suas condições dinâmicas permite melhor compreensão dos fenômenos observados e uma tomada de decisões mais precisa e ágil, quando alguém está acompanhando comentários sobre um assunto, por exemplo.

O Modelo Orbital proposto e sua visualização permitem uma compreensão clara e dinâmica das relações entre os termos de interesse em um grande volume de informações textuais.

Uma das principais limitações envolvidas nesta abordagem de modelagem pode ser atribuída à Folksonomia em si mesma, porque se baseia na opinião de usuários que podem não ser os mais adequados para sugerir etiquetas. Além do amadorismo dos usuários quando se trata de comentar sobre o assunto a ser monitorado, as etiquetas podem ser incorretamente escolhidas ou organizadas, o que pode refletir certo preconceito por parte dos usuários que participaram dos processos de etiquetagem. Assim, uma análise preliminar e filtragem de usuários com um perfil (ou *background*) mais adequado para comentar

sobre o assunto específico pode contribuir para o aumento da qualidade de categorização e do resultado final.

A forma gráfica proposta neste Modelo Orbital pode ser estendida para a análise de dados de outras mídias sociais como o Facebook, por isso, a ferramenta deve ser adaptada para a coleta de dados sobre ele. Além disso, o modelo pode ser utilizado para a análise de outros conteúdos diferentes, como vídeos, imagens etc., utilizados nas mídias sociais, melhorando as observações de como as mídias mais populares estão sendo escolhidas para a discussão de determinados assuntos.

O estudo deve ser estendido para a análise de outros temas para verificar suas possíveis variações na determinação do valor de K com outros dados e em outras línguas.

Quanto à aplicação dos resultados, esta forma de visualização de dados pode ser utilizada em áreas específicas, tais como a comercialização de produtos ou campanhas políticas, a fim de se acompanhar como as discussões estão ocorrendo nas mídias sociais, levando em consideração que a busca por ferramentas visuais tem sido uma demanda atual (Karpinsk, 2009; Shearman, 2012).

O trabalho futuro centrar-se-á na escolha de dados sobre as próximas campanhas políticas brasileiras e como o Modelo Orbital poderá ajudar a entender a opinião dos usuários sobre os candidatos ao longo do tempo.

Referências

Becker, H., Naaman, M. & Gravano, L. (2011). "Beyond trending topics: real-world event identification on Twitter". In: *Proceedings of Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, p. 438-441.

Boguta, K. (2009). "Evolution of a revolution: visualizing millions of Iran tweets". ReadWriteWeb. Retrieved June, 2012, from http://www.readwriteweb.com/archives/evolution_revolution_visualizing_millions_iran_tweets.php

Boyd, D. & Crawford, K. (2012). "Critical questions for big data". Information, Communication & Society, 15 (5), 662–679.

Cantadora, I., Konstas, I. & Josec, J. M. (2011). "Categorising social tags to improve folksonomia-based recommendations". Web Semantics: Science, Services and Agents on the World Wide Web, 9, 1–15.

Chen, I.-X. & Yang, C.-Z. (2010). "Visualization of social networks". In: B. Furht (ed.). *Handbook of Social Network Technologies and Applications*: Springer Science+Business Media.

Cheong, M. & Lee, V. C. S. (2010). "Twitmographics: learning the emergent properties of the Twitter community". In: *From Sociology to Computing in Social Networks*, 323-342: Springer-Verlag.

_____. (2011). "A microblogging-based approach to terrorism informatics: exploration and chronicling civilian sentiment and response to terrorism events via Twitter". Information Systems Frontiers, 13, 45–59.

Fleishman, J. (2009). "Mideast hanging on every text and tweet from Iran". Los Angeles Times. Retrieved May, 2012, from http://articles.latimes.com/2009/jun/17/world/fg-iran-image17

Gilbert, E., Karahalios, K. & Sandvig, C. (2010). "The network in the garden: designing social media for rural life". American Behavioral Scientist, 53(9), 1367–1388: SAGE Publications.

Go, A., Bhayani, R. & Huang, L. (2009). "Twitter sentiment classification using distant supervision". Stanford University. Retrieved June, 2012, from http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf

Goolsby, R. (2009). "Lifting elephants: Twitter and blogging in a global perspective". In: *Social computing and behavioral modeling*: Springer-Verlag.

Gupta, M., Li, R., Yin, Z. & Han, J. (2011). "An overview of social tagging and applications". In: Aggarwal , C. C. (ed.). *Social Network Data Analytics*: Springer Science+Business Media.

Highfield, T., Kirchhoff, L. & Nicolai, T. (2011). "Challenges of tracking topical discussion networks online". Social Science Computer Review, 29(3), 340-353.

Himelboim, I., Hansen, D. & Bowser, A. (2012). "Playing in the same Twitter network. Information", Communication & Society. DOI: 0.1080/1369118X.2012.706316.

Jansen, B. J., Zhang, M., Sobel, K. & Chowdury, A. (2009). "Micro-blogging as online word of mouth branding". Conference on Human Factors in Computing Systems (CHI 2009) (pp. 3859-3864). Boston, MA: ACM.

Java, A., Song, X., Finin, T. & Tseng, B. (2007). "Why we Twitter: understanding microblogging usage and communities". In: *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp. 56-65, 2007. Proceedings of the 9th WEBKDD and 1st SNA-KDD 2007, (pp.118-138): San Jose, CA.

Jin, Y., Li, R., Wen, K., Gu, X. & Xiao, F. (2011). "Topic-based ranking in folksonomia via probabilistic model". Artificial Intelligence Review, 36, 139–151.

Jungherr, A. (2009). "The DigiActive guide to Twitter for activism". Retrieved May, 2012, from http://tinyurl.com/6tr2pfp

Kakali, C. & Papatheodorou, C. (2010). "Exploitation of folksonomies in subject analysis". Library & Information Science Research, 32, 192–202.

Karpinski, R. (2009). *Twitter tools*. B to B, vol. 94, n. (2), 2009, pp. 15-20.

Kawano, Y., Kishimoto, Y. & Yonekura, T. (2011). "A Prototype of Attention Simulator on Twitter". International Conference on Network-Based Information Systems. IEEE Computer Society.

Kim, H.-N., Rawashdeh, M., Alghamdi, A. & El Saddik, A. (2012). "Folksonomia-based personalized search and ranking in social media services". Information Systems, 37, 61–76.

Lin, J. & Dyer, C. (2010). *Data-Intensive Text Processing with MapReduce*. University of Maryland: College Park.

Liu, B. (2010). "Sentiment analysis and subjectivity". In: Indurkhya, N & Damerau F. J. (Eds). *Handbook of Natural Language Processing* (pp. 627-666). Boca Raton, FL: Chapman & Hall/CRC.

López-Juárez, P. & Olivas, J. A. (2011). "Intentional tags in folksonomia based ranking systems". *The 2011 World Congress in Computer Science, Computer Engineering*, and Applied Computing. Retrieved June, 2012, from http://world-comp.org/p2011/ICA5049.pdf

Luo, X., Ouyang, Y. & Xiong, Z. (2012). "Improving neighborhood based Collaborative Filtering via integrated folksonomia information". Pattern Recognition Letters, 33, 263–270.

MerlesWorld. A Beginner's Guide to Using & Marketing with Twitter. Retrieved Disponível em: June, 2012, from de http://www.merlesworld.com/e-books/Twitter_Fast_Start.pdf. Acesso em junho de 2012.

Naaman, M., Becker, H. & Gravano, L. (2011). "Hip and trendy: characterizing emerging trends on Twitter". Journal of the American Society for Information Science and Technology, 62(5), 902–918.

Nicholls, J. (2012). "Everyday, everywhere: alcohol marketing and social media - current trends". Alcohol and Alcoholism. DOI:10.1093/alcalc/ags043.

Pang, B. & Lee, L. (2008). "Opinion mining and sentiment analysis". Foundations and Trends in Information Retrieval, 2 (1), 1–135.

Ran, Z. & Erpeng, J. (2011). "Folksonomia - based library information organization in China". International Conference on Business Management and Electronic Information (BMEI), 5, 270-272.

Romero, D. M., Meeder, B. & Kleinberg, J. (2011). "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter". International World Wide Web Conference (WWW 2011): India.

Schmitz,C., Hotho, A., Jaschke, R. & Stumme, G. (2006). "Mining association rules in folksonomies". In: *Data Science and Classification:* Proceedings of the 10th IFCS Conference (pp. 261-270), Studies in Classification, Data Analysis, and Knowledge Organization: Springer.

Shearman, S. (2012). "Twitter's branded venture". Marketing, 10-11.

Skold, M. (2008). *Social network visualization*. Master Thesis. Royal Institute of Technology: Sweden.

Thelwall, M., Buckley, K. & Paltoglou, G. (2011). "Sentiment in Twitter events". Journal of the American Society for Information Science and Technology, 62(2), 406-418.

Tsytsauru, M. & Palpanas, T. (2012). "Survey on mining subjective data on the web". Data Mining and Knowledge Discovery, 24(3), 478-514.

Wilson, V. "Research methods: content analysis". Evidence Based Library and Information Practice, 6(4), 177-179.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M. & Liu, B. (2011). "Combining lexicon based and learning-based methods for twitter sentiment analysis". Technical Report HPL-2011-89: HP.